AD-A017 881

ERROR IN DECISION ANALYSIS: HOW TO CREATE THE
POSSIBILITY OF LARGE LOSSES BY USING DOMINATED
STRATEGIES

Detlof von Winterfeldt, et al

University of Southern California

Prepared for:

Office of Naval Research
Advanced Research Projects Agency

1 April 1975

339102

001597-1-T

AD A017881

social science
research institute

UNIVERSITY OF SOUTHERN CALIFORNIA

Technical Report

Error in Decision Analysis: How to Create
The Possibility of Large Losses By Using
Dominated Strategies

Detlof von Winterfeldt
Ward Edwards

Sponsored by:

Advanced Research Projects Agency
Department of Defense

Monitored by:

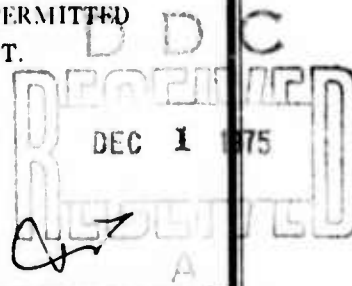Engineering Psychology Programs
Office of Naval Research
Contract No. N00014-67-A-0269-0034
NR 197-021

April 1975

DEC 1 1975

SSRI Research Report 75-4

60

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER OO159-1-T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Error in Decision Analysis: How to Create the Possibility of Large Losses by Using Dominated Strategies		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER None
7. AUTHOR(s) Detlof von Winterfeldt and Ward Edwards		8. CONTRACT OR GRANT NUMBER(s) N00014-67-A-0269-0034
9. PERFORMING ORGANIZATION NAME AND ADDRESS Social Science Research Institute University of Southern California Los Angeles, California 90007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 197-021 ARPA Order No. 2105
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209		12. REPORT DATE 1 April 1975
		13. NUMBER OF PAGES 60
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Engineering Psychology Programs Office of Naval Research Arlington, Virginia 22217		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Decision Analysis	Sensitivity
Dominance	Information Use
Admissibility	Value of Information
Flat Maxima	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report examines some concepts, sources, and possible consequences of error in decision analysis. Recent articles on the possibilities for error in decision analysis showed that under some relatively mild assumptions deviations from optimal decision strategies or from optimal model parameters will lead only to minor losses in expected value. This "flat maximum" property of decision analytic models applies, however, only to admissible decisions. By inadvertently selecting a dominated (inadmissible) decision, the decision
(over)

DD FORM 1473 1 JAN 73   EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

(Cont.)

maker creates the possibility for large expected losses. Usually dominance can be recognized and losses can be avoided by elimination of dominated decisions. Unfortunatley, for a large class of errors the discovery of dominance is difficult if not impossible. These errors consist of failing to use information or using it inappropriately in decision strategies. The main point this report makes is that such errors can, and typically will, lead to dominated strategies, and so can lead to substantial expected losses.

ERROR IN DECISION ANALYSIS: HOW TO CREATE THE
POSSIBILITY OF LARGE LOSSES BY USING DOMINATED
STRATEGIES

Technical Report

1 April 1975

Detlof von Winterfeldt and Ward Edwards

Social Science Research Institute
University of Southern California

Decision analysis is an integrated collection of formal and behavioral tools for solving complex decision problems. Its essential steps are to analyze the decision problem into a formal structure (most typically a decision tree; see Raiffa, 1968, or Brown, Kahr, and Peterson, 1974), to elicit from the decision maker or his surrogate(s) some relevant numbers of two types, probabilities and values, and then to apply suitable formal arithmetic that permits calculation of the expected value of each course of action under consideration. The result may be a decision in favor of the act with the highest expected utility; more often, it is an exploration of alternative formal structures and, within each structure, of numbers that the decision maker might have estimated instead of those he did estimate. These exploration processes are jointly called <u>sensitivity analysis</u>; most final decisions emerging from decision analysis should be, and are, supported by sensitivity analyses. Although much mathematical and behavioral sophistication is required for intelligent sensitivity analysis, it is also relatively unsystematic; sensitivity analysis is more art than science.

<u>The Decision Analyst's Cop-Out</u>. Can a decision analysis be wrong? Throughout this paper we ignore the possibilities of arithmetical error, misunderstandings between analyst and decision maker, insufficiently deep deliberation by either, and similar potential sources of errors large and small that seem to offer little scope for formal analysis. The question we are asking is: if the decision analyst and decision maker each conscientiously and diligently performs his part in a decision analysis, and no stupid or inadvertent errors are made by either, to what extent and for what reasons can the actions resulting from it, be wrong? While the notion of "stupid

1

or inadvertent error" is a bit ill-defined, this definition is sufficiently precise to work with for the moment.

Any other modeling process can be wrong as a basis for action in either or both of two ways: the model may be wrong, in the sense of being either misleading or too crude a representation of the phenomenon modeled; or the data may be wrong, in any of a variety of ways. Presumably the same two possibilities exist for decision analysis.

Yet the formal literature of decision analysis is remarkably quiet about these possibilities. About the possibility that the model may be wrong, decision analysis have maintained virtually complete silence (though a recent unpublished paper by Brown may be an exception). About the possibility that the data may be wrong, decision analysts have been more helpful; they have offered a few consistency rules, mostly those of formal probability theory, that judged numbers obtained from decision makers or their surrogates should obey.

These rules enforce consistency, not good sense; they are exactly as usable by the inhabitant of the local asylum who, believing that he is Napoleon, wishes to plan his reconquest of Europe as by the breakfast food magnate who wishes to plan the marketing strategy that will put a box of Munchy-Scrunchy-Wunchies on each of a billion breakfast tables within the next year.

This silence about wrong models and near-silence about wrong data is not accidental. It arises from a line of reasoning that, though in a sense we subscribe to it, we shall call by a pejorative, provocative name: the Decision Analyst's Cop-Out.

The Decision Analyst's Cop-Out grows out of a set of methodological principles:

1. Values are inherently subjective; and the values that should be maximized in making a decision are those of the decision maker. (Throughout this paper we shall use the words "value" and "utility" interchangeably to mean subjective value.)

2. Probabilities are inherently personal, in the sense that they describe orderly opinions about the likeliness of uncertain events; and the opinions that should be used in making a decision are those of the decision maker.

3. The function of decision analysis and of the decision analyst is to help the decision maker to make wise decisions by helping him to understand the ideas of decision analysis, by helping him to model his problem in decision-analytic form, by helping to translate his values and probabilities into explicit, numerical form, and by checking for logical consistency within and between opinions, values, and actions chosen.

4. A wise decision, in the contexts to which this paper applies, is one that maximizes expected utility. An implication of this definition is that wise decisions can lead to unfortunate or even disastrous outcomes. Every decision under uncertainty is in effect a bet, and any bet can be won or lost (intermediate outcomes are usually also possible).

5. By the use of suitable elicitation techniques, the decision analyst can elicit from the decision maker an accurate representation of his model of his decision problem and numbers that accurately represent his values and probabilities.

While no decision analyst would, we trust, accept these principles in quite the blunt form we have used above, the typical modifications would be of tone, emphasis, and softening of claims for precision, rather than of

substance. If these principles are a caricature, they are an easily recog-
nizable one.

Now we can state the Decision Analyst's Cop-Out explicitly: given the
above principles a decision analysis cannot be wrong.

The logic is obvious enough. The model of the problem is the decision
maker's, not the analyst's. The same is true of values and probabilities. If
the analyst has been sufficiently assiduous at his elicitation tasks, then all
three will be suitable representations of the inside of the decision maker's
head--and that is the only test of representation they are required to meet.
If the actions chosen are consistent with the results of the analysis, and the
numbers obey the appropriate consistency rules, it makes no difference whether
others would consider these actions and numbers wise or foolish, or whether
they lead to happy or unhappy results; they maximize the expected utility of
the decision maker, and that is all they are called on to do.

In short, the only explicit test of the adequacy of a decision analysis
is obedience to internal consistency rules. Others may criticise models,
numbers, or both, but such criticisms are in principle irrelevant (though in
practice every analyst would take them seriously indeed).

Given this line of reasoning, why would a decision analyst undertake a
sensitivity analysis at all? There seem to be two answers. One is intel-
lectual curiosity. The other is that no decision analyst really believes
Principle 5. He is never entirely sure that he has elicited exactly the right
model structure, or exactly the right numbers, from the decision maker, and so
he wishes to reassure himself that minor errors of these kinds would make
little difference to the outcome. Virtually always, such reassurance is
provided.

<u>Can we escape from the Decision Analyst's Cop-Out?</u> To escape from the
intellectual trap outlined above, one must reject or modify one or more of the
five methodological principles from which it derives. Principle 1 dates from
the Greeks and is a cornerstone of disciplines ranging from ethics (where
everyone seems agreed that it is wrong, but few agree about why) to economics.
Principle 2 is the fundamental tenet of the personalist Bayesian school of
probability. While it has provoked much modern debate (see for example
Savage, 1954; Edwards, Lindman, and Savage, 1963; and references cited there-
in), we believe that the Bayesians have clearly won the argument. No intel-
lectually viable set of identification rules for probabilities alternative to
those implied by Principle 2 has yet been proposed, so far as we know.

Principle 3 is of relatively minor importance; it probably is no more
than a consequence of Principle 1. We included it in our list more to give
the intellectual flavor of the Decision Analyst's Cop-Out than because of any
logical role it plays in that line of reasoning.

Principle 4 seems to us beyond question, given two conditions that are
assumed throughout this paper. One is that the decision problem is a Game
against Nature, which simply means a decision in which the concept of a
hostile opponent whose actions depend on those of the decision maker plays no
significant role. The other is that the stakes are not so large as to include
the possibility of ruin or quasi-ruin. (Actually, both restrictions can be
removed by appropriate interpretations of the notions of utility and proba-
bility, but the topic is complex and irrelevant to the purpose of this paper.)

Principle 5 is obviously the most dubious one. As we have already said,
no one would believe it literally, except perhaps a radical behaviorist--and
radical behaviorists who are decision analysts are rare indeed. If the

decision maker's "true" value and probability are u and p respectively, an elicitation procedure may well cause him to estimate $u' \neq u$ and $p' \neq p$. Both u' and p' may pass all relevant internal consistency tests. If the differences are large, the analyst would expect to discover them by various checking procedures, but if they are small, he might not. Of course, intuition suggests that small differences in such elicited numbers are unlikely to lead to large differences in the expected utilities that are the near-final outputs of a decision analysis--and we have recently proven exactly that under very general conditions (von Winterfeldt and Edwards, 1973b). The maxima of decision analysis are flat. Relatively substantial deviations between "true" parameters and those used in making a decision-analytic calculation will typically produce relatively minor reductions, if any, in expected utility of the action chosen from the expected utility that would have been produced by the best criterion had the "true" parameters been used.

But Principle 5 covers not only values and probabilities, but also models. Here the chances for error seem to be, and are, much greater. As a matter of realism, we are skeptical about the idea that decision makers have explicit models of their problem in their heads, waiting to be elicited. Instead, they have ideas, of widely varying degrees of coherence, intelligibility, and appropriateness, about the nature of their problem. The decision analyst may indeed elicit these ideas, but typically it is he, not the decision maker, who formulates from them and from other available information an explicit, well-defined model of the decision problem. And as a matter both of realism and of good sense, he is more likely to worry about whether the model fits the problem than about whether it fits the decision maker's ideas about that problem--though obviously both kinds of fit are important. (This, of course, violates Principle 3.)

This paper is about a rather subtle set of errors that can occur in modeling decision problems, and that can lead to very large errors in the resulting analysis. In concept these errors are not subtle: they consist of failing to use information, or using it inappropriately, to modify probabilities. The main point we make is that such errors can, and ordinarily will, lead to use of dominated strategies, and so can lead to grossly suboptimal actions. The reason why we consider such errors subtle is that they are relatively difficult for the decision analyst to discover. He often must rely on the decision maker to specify what information is relevant to a decision, and what that information means. Decision makers, unfortunately, often ignore relevant information, or use it in grossly inappropriate ways.

In what follows we will be considering only cases in which selection of an information source and quality of information processing are not explicitly modeled within the decision analysis. Only in the absence of such modeling can unrecognizable uses of dominated strategies occur.

Why, then, would anyone perform a decision analysis in which choice of information source and information processing technique were left out? Often, they are left out because the information required to model them is unavailable. This is especially likely if the information comes from or is processed by a human being. Indeed, in the example of Fryback's (1974) study that led us to consider this problem, one conclusion of the study was that different radiologists differ greatly in their ability to extract information from a particular kind of radiograph. After the study, one might well know that it was far better to ask radiologist X to read the film than to ask radiologist Y--but before the study, or in its absence, how could one know that Y was sufficiently inferior to X so that any strategy based

on Y's readings would be dominated by any admissible strategy based on X's? In the absence of this kind of information, it is easy to have bad luck in choice of an information source without knowing it. In this paper, we have called such bad luck dominance and have chosen to omit the selection of an information source from the formal analysis. It would instead be possible to draw the decision tree in such a fashion as to include the choice between X and Y as one of the decisions (or, more realistically, as one of the random events controlled by Nature). For the analysis, little would be gained by doing so in the absence of any information about the relative skills of X and Y. However, if that were included as a specific choice, then a strategy that includes having Y read the radiogram would appear as bad luck, rather than dominance. (The extreme case of this argument is decision making under certainty, in which all strategies but one are dominated.)

Our point, then, is fundamentally that poor information, or poorly processed information, can lead to severely painful consequences. Whether one calls this dominance or bad luck is fairly irrelevant. Either way, it is a major source of loss in decision analyses, and a major exception to the general principle of flat maxima that applies elsewhere in such analyses.

We do not mean to imply that every decision analysis has built into it the possibility of major loss, avertable only by great expertise and perfect information processing. Indeed, in many decision environments cost of information is positively related to its benefit. Even if the information acquisition and processing aspects of the problem are not modeled, this trade-off may mean that the consequences of inferior information are offset by the fact that poor information is cheaper than good information. Of course this is not always true; radiologists X and Y charge the same fees.

Since this paper is about error, it is also about sensitivity analysis, the decision analyst's most important tool for discovering and correcting errors. We believe that some strategies for allocating sensitivity analysis effort, and even some specific tools, grow out of the arguments to be presented --though by no means enough to change sensitivity analysis from art to science.

Social Decision Making. This preview of error in decision analysis would be severely incomplete if it did not at least mention the most serious source of error of them all: The fact that for many decisions the concept of "the decision maker" that is embodied in Principles 1-5 is just not applicable. Most decisions affect many people, not just one. Some are made by many people; perhaps working cooperatively, perhaps not. Even a single decision maker is likely to explore the values and probabilities of others before making a socially important decision. Often, the most important kind of service he might want from a decision analyst (a service, alas, that may not be available for lack of necessary conceptual tools) is that of reconciling or otherwise dealing with conflicting values and/or probabilities.

Re-examine for a moment Principles 1-5 with some major social decision, such as imposition of gasoline rationing, in mind. The relevant values are those of everyone affected; some undefined amalgam of them should presumably be maximized, but no one knows even how to define, much less how to calculate, that amalgam. While virtually all of those affected will have opinions about the questions of uncertainty that bear on the decision (e.g., will there be another war between Arabs and Israelis), most of those opinions will be worth-less as a basis for action, except insofar as they constrain or bias the action options or influence the relevant values. The opinions worth considering will be those of a relatively small community of experts, most of whom have studied the problem for years. The decision maker, if there is one,

is seldom a member of that community; indeed its members usually have information about the topic that the decision maker cannot hope even to read, much less to understand. If the community of experts agreed, the decision maker might be able simply to treat their opinion as his own. But members of such communities of experts typically disagree. Again, some aggregation process is needed, but no one knows what it is. Finally, the decision analyst, if he is to be a responsible member of the decision-making team or organization, cannot limit his role to effective use of the formal tools of his trade. His job is the same as that of each other member of the team: by hook or by crook to see to it that the wisest available actions get taken. Obedience to consistency rules is far from enough!

Many of the issues raised in the preceding paragraph are under current study. A few rudimentary tools exist; more can be foreseen. In the absence of a well-developed theory and technology of social decision making, we cannot hope to analyze sources of error in that technology. This paper consequently applies only the concepts behind the technology of individual decision making to the study of the error-producing potential of that technology. That is our cop-out.

Large Losses in Decision Analysis

The Puzzle of Flat Maxima and Large Losses. We recently discussed some
of the possibilities for error in decision analysis in our treatment of flat
maxima (v. Winterfeldt and Edwards, 1973a and b). We showed that under some
relatively mild assumptions, so we thought, suboptimal probabilities, values,
or model parameters in a decision analysis will lead to only minor losses in
expected utility. Loosely speaking, once a decision problem has been properly
formulated and once grossly inappropriate (by which we mean dominated or
cardinally dominated) strategies are eliminated from consideration, the mathe-
matical properties of the usual maximization processes impose severe restric-
tions on the functions used to evaluate available action or decision strategy
alternatives. These restrictions almost always result in rather flat func-
tions in the "neighborhood" of the optimal decision or decision strategy.
It takes large errors of the numbers entering into the analysis to lead to
choices of actions or strategies outside of that neighborhood; and within it,
reductions of expected value from the optimal expected value are quite small.
A 10 percent reduction would be unusually large.

Proper scoring rules (see Murphy and Winkler, 1970), signal detection
tasks (see Green, 1967), and decisions about optimal sample sizes (see
Schlaifer, 1969) are well known to have this sort of flat-maximum property;
and although we know of no convenient reference, it has been common knowledge
among decision analysts that changing model parameters often produces only
minor changes in the result of a decision analysis.

Nevertheless, real world decision making reminds us that substantial losses can and do occur. Individual instances can be attributed to bad luck; good decisions can of course lead to bad outcomes. But such bad luck should average out as instances accumulate. So we were especially impressed by Fryback's (1974) finding that in a real-world medical decision problem, although the functions showing the relation between size of error in decision strategy and resulting loss in expected utility were even flatter than we might have expected a priori in a long series of cases the doctors were actually obtaining only a little more than 50 percent of the expected utility obtainable by the decision-theoretically optimal procedure.

On reflection, we realized that our flat-maximum analysis had failed to deal with two important facts. One is that real decisions are typically made without proper prior decision-analytic structuring, and in particular without prior elimination of grossly inappropriate decisions or strategies. The other is that the flat-maximum ideas apply only to the decision making part of a decision analysis, not to the information processing part. Neglect or inefficient use of information can in effect create dominated strategies, not recognizable as such from inspection of payoff matrices or decision trees, and can make these dominated strategies seem optimal.

Dominance and the intimately related concept of admissibility are the key concepts in these instances of large losses. In the following sections we first define, classify, and relate concepts of dominance and admissibility. Then we show that inefficient use of information leads to dominated strategies and use of dominated strategies leads to substantial losses. Finally, we examine implications for the design of decision analysis and for decision theoretic thinking in general.

Our discussion is based mostly on some well known theorems of statistical decision theory. If not explicitly stated and proved here, they can be found in various of the following sources: Blackwell and Girshick, 1954; Lehman, 1959; Raiffa and Schlaifer, 1961; Ferguson, 1967; and DeGroot, 1970.

<u>Definitions and Assumptions</u>. The first step in an analytic treatment of a complex decision problem is to structure the problem in the form of a decision tree or some equivalent description. A decision tree consists of decision nodes and chance nodes. At each decision node the decision maker decides between alternative courses of action. At each chance node a probabilistic process determines which state of nature or which value of an observation variable obtains.

It is useful to define an exhaustive set of the three kinds of acts that can be available to a decision maker at a decision node:

1. Acts that directly produce a riskless (but possibly multi-
    attributed or time variable) outcome. We will call the set
    of outcomes A with typical elements a, b, c, ...

2. Acts that will result in some outcome element from A, which
    element depends on which state of nature obtains. Such
    acts are also called gambles, and we will denote the set
    of gambles as G with typical elemnents $\underline{a}$, $\underline{b}$, $\underline{c}$, ...

3. Acts that first result in the acquisition of an observation.
    We will call the set of possible observations X. Upon obser-
    vation of a particular value x_j in X a preselected function
    (decision rule) determines which element of the set G to choose.
    Such acts are also called decision functions, and we will de-
    note the set of decision functions as D with typical elements
    $\underline{\underline{a}}$, $\underline{\underline{b}}$, $\underline{\underline{c}}$, ...

At each chance node either of two random processes may occur:

1. A process which selects from the states of nature S a particular element S_i.

2. A process which selects from the set X of observations a particular element x_j.

(Some of the following definitions require the sets X and S to be finite. This assumption will be made from now on unless especially noted.)

A simplified decision tree in which all three types of acts are represented is sketched in Figure 1.

Insert Figure 1 about here

The tree structure suggests the following representation of outcomes, gambles, and decision functions as real numbers, vectors, and matrices:

1. <u>Outcomes</u>: An element a in A is called an outcome. Elements in A are evaluated according to some real valued function $U : A \rightarrow R$ which preserves the order of preferences among outcomes. For simplicity of notation, we will assume that A is real valued and that $U(a) = a$.

2. <u>Gambles</u>: An element in G is called a gamble. A gamble is an n-element vector of elements in A:

$\underline{g} = (g_1, g_2, \ldots, g_i, \ldots, g_n); \; g_i \epsilon$ A, where g_i is the outcome which the decision maker receives if the i-th state of nature obtains. We assume that the expected value of such gambles preserves their preference order. The expected value is defined as

$$EV \; (\underline{g}, \; f) = \sum_{i=1}^{n} f(S_i)g_i \tag{1}$$

where f is the probability distribution over the states of nature.

   3. <u>Decision functions</u>: An element $\underline{\underline{d}}$ in D is called a decision function. A decision function $\underline{\underline{d}} : X \rightarrow G$ is a function from the observation variable into the set of gambles. If X and S are finite, a decision function can be described by an n x m matrix of elements in A in which the row $\underline{\underline{d}}(x_j)$ is an element in G which the decision maker selects if the observation variable has value x_j.

$$
\underline{\underline{d}} =
\begin{bmatrix}
d_{11} & d_{21} & \cdots & d_{i1} & \cdots & d_{n1} \\
\vdots & \vdots & & \vdots & & \vdots \\
d_{1j} & d_{2j} & \cdots & d_{ij} & \cdots & d_{nj} \\
\vdots & \vdots & & \vdots & & \vdots \\
d_{1m} & d_{2m} & \cdots & d_{im} & \cdots & d_{nm}
\end{bmatrix}
=
\begin{bmatrix}
\underline{\underline{d}}(x_1) \\
\vdots \\
\underline{\underline{d}}(x_j) \\
\vdots \\
\underline{\underline{d}}(x_m)
\end{bmatrix}
$$

$$d_{ij} \epsilon A , \quad \underline{\underline{d}}(x_j) \epsilon G$$

We assume that the expected value preserves the preference order over decision functions. The expected value of a decision function is defined as

$$
EV (\underline{\underline{d}},f) = \sum_{i=1}^{n} f(S_i) \sum_{j=1}^{m} g(x_j|S_i) d_{ij}
\tag{2}
$$

where f is the prior distribution over the states of nature, and $g(X|S_i)$ are the respective conditional distributions over the observation variable X.

   The preceding discussion sounds as though observing, and processing the resulting information, were assumed to be costless. They are, but with no loss of generality. Given additive utilities, which are assumed throughout

this paper, the utility cost of an observing-and-processing procedure can be subtracted from the utility payoffs associated with each possible ultimate outcome of that procedure. If that has been done, no further attention need be paid to the cost of that procedure in the analysis.

To illustrate the concepts of outcomes, gambles, and decision functions, consider certain amounts of money as outcomes. Let

$$A = \{ \$1, \$2, \$3, \$4, \$5, \$6, \$7 \}.$$

Let two mutually exclusive and exhaustive states of nature, S_1 and S_2, determine which outcome the decision maker will receive. A gamble is then of the form $\underline{g} = (a,b)$, where the decision maker receives amount a if state S_1 obtains, b otherwise. Assume that the decision maker has the option to select one of the following gambles:

$$G = \{(2,3); (2,6); (3,5); (6,4); (7,3); (1,1)\}.$$

To explain the idea of a decision function, assume that before choosing a gamble the decision maker can observe a random variable X which can obtain either of two values x_1 or x_2. The probability distribution over X depends on the state of nature S_i. Let

$$g(x_1|S_1) = 1-g(x_2|S_1) \text{ and}$$

$$h(x_1|S_2) = 1-h(x_2|S_2).$$

describe the two distributions.

The following are examples of the 36 possible decision functions among which the decision maker can select:

$$\underline{d}_1 = \begin{bmatrix} 2, & 3 \\ 2, & 3 \end{bmatrix}$$

$$\underline{\underline{d}}_2 = \begin{bmatrix} 3, & 5 \\ 7, & 3 \end{bmatrix}$$

$$\underline{\underline{d}}_3 = \begin{bmatrix} 7, & 3 \\ 2, & 6 \end{bmatrix}$$

For example, in $\underline{\underline{d}}_3$, the decision maker would select the gamble $(7, 3)$ if x_1 occurs, and gamble $(2, 6)$ if x_2 occurs. Thereafter he will receive a specific amount in A depending on which state of nature is true.

Now assume that the decision problem is specified by three possible courses of action :

1) take $3 for sure;

2) select any of the six gambles in G;

3) select any of the 36 decision functions in D.

The optimal course of action depends, of course, on the prior distribution f over the states of nature, and on the conditional distributions g and h. Given the knowledge of these distributions, the decision maker can determine the expected value of each course of action. As an expected value maximizer he should select the course of action that guarantees him the highest expected value. For example, let

$$f(S_1) = \frac{1}{2} \; ; \; f(S_2) = \frac{1}{2} \; ;$$

$$g(x_1|S_1) = \frac{1}{3} \; ; \; g(x_2|S_1) = \frac{2}{3} \; ;$$

$$h(x_1|S_2) = \frac{2}{3} \; ; \; h(x_2|S_2) = \frac{1}{3} \; .$$

The expected value of the first course of action is $3, independently of the distributions f, g, and h. The expected value of the second course of action is that of the gamble in G which has maximal expected value. In our example $(6, 4)$ and $(7, 3)$ have the maximal EV with

$$EV\ (6,\ 4) = \frac{1}{2}\ (6+4) = 5 = EV\ (7,\ 3) = \frac{1}{2}\ (7+3).$$

Among the 36 decision functions two maximize expected value. From the general expected value formula for decision functions (see p. 13), it follows that the expected value of a decision function in our example is

$$EV(\underline{d}) = \frac{1}{2}\ [\frac{1}{3}\ (d_{11}) + \frac{2}{3}\ (d_{12})] + \frac{1}{2}\ [\frac{2}{3}\ (d_{21}) + \frac{1}{3}\ (d_{22})],$$

an expression which is jointly maximized by the two decision functions

$$\underline{d}_1 \ = \ \begin{bmatrix} 2 & 6 \\ 7 & 3 \end{bmatrix}$$

and

$$\underline{d}_2 \ = \ \begin{bmatrix} 6 & 4 \\ 7 & 3 \end{bmatrix}$$

with

$$EV(\underline{d}_1) = \frac{1}{2}\ [\ \frac{1}{3}\ (2) + \frac{2}{3}\ (6)\ ] + \frac{1}{2}\ [\ \frac{2}{3}\ (7) + \frac{1}{3}\ (3)\ ] = 5.16667 \text{ and}$$

$$EV(\underline{d}_2) = \frac{1}{2}\ [\ \frac{1}{3}\ (6) + \frac{2}{3}\ (4)\ ] + \frac{1}{2}\ [\ \frac{2}{3}\ (7) + \frac{1}{3}\ (3)\ ] = 5.16667.$$

Therefore, the decision maker should select the third course of action, since the best decision functions have a higher expected value than the best gamble, which in turn has a higher expected value than the sure amount \$3. Of course, this conclusion holds only for the specified distributions f, g, and h.

However, even in the absence of any knowledge about f, the decision maker can make some evaluation of the decision alternatives by assessing the value or the expected value of gambles or decision functions conditional on the assumption that the true state of nature is S_i. We will call these conditional

evaluations the conditional values of gambles and decision functions, and we define them formally as:

Definition: Conditional values. The i-th conditional value of a gamble $\underline{g} \varepsilon G$ is defined as

$$CV_i (\underline{g}) = g_i \quad ; \tag{3}$$

the i-th conditional value of a decision function $\underline{\underline{d}} \varepsilon D$ is defined as

$$CV_i(\underline{\underline{d}}) = \sum_{j=1}^{m} g(x_j | S_i) d_{ij} \tag{4}$$

We can now describe gambles and decision functions as vectors of conditional values. We will call the set of conditional value vectors of gambles $\underline{G}$ and the set of conditional value vectors of decision functions $\underline{D}$. Sometimes we will consider the set of all possible probability mixtures of gambles, denoted as $\tilde{G}$, or of decision functions, denoted as $\tilde{D}$. The set of conditional value vectors of $\tilde{G}$ and $\tilde{D}$ will be written as $\underline{\tilde{G}}$ and $\underline{\tilde{D}}$, respectively.

Figures 2 and 3 are plots of conditional values of gambles and decision functions in our examples. The points in these plots constitute $\underline{G}$ and $\underline{D}$, respectively. The conditional value vectors of probability mixtures of gambles and decision functions, $\underline{\tilde{G}}$ and $\underline{\tilde{D}}$, lie in the closed and convex region defined by the points in $\underline{G}$ and $\underline{D}$. Conversely, any point within that region is a conditional value vector for some probability mixture of gambles or decision functions (see Ferguson, 1967). In that sense, the shaded areas in Figures 2 and 3 describe $\underline{\tilde{G}}$ and $\underline{\tilde{D}}$, respectively; that is, they define the set of all possible points equivalent in expected value to probability mixtures of $\underline{G}$ and $\underline{D}$. The circled points in Figure 3 show where the conditional value vectors of gambles lie in $\underline{\tilde{D}}$.

Insert Figures 2 and 3 about here

The following implications should be obvious (see Blackwell and Girshick, 1954):

Lemma 1
$$\underline{G} \subseteq \tilde{\underline{G}}$$
$$\underline{D} \subseteq \tilde{\underline{D}}$$
$$\underline{G} \subseteq \underline{D}$$
$$\tilde{\underline{G}} \subseteq \tilde{\underline{D}}$$

The last result follows by letting $\underline{d}(x_j) = \underline{g}$ for all j.

The following definitions are stated for decision functions only, but they apply - mutatis mutandis - to gambles also.

Definition: <u>Ordinal dominance</u>. A decision function $\underline{e} \varepsilon D$ is said to be ordinally dominated, if there exists another decision function $\underline{d} \varepsilon D$ such that

$$CV_i \ (\underline{d}) \geq CV_i \ (\underline{e}) \quad \text{for all i;}$$

$$CV_i \ (\underline{d}) > CV_i \ (\underline{e}) \quad \text{for at least one i.}$$

A stronger and more useful definition is the following.

Definition: <u>Cardinal dominance</u>. A decision function $\underline{e} \varepsilon D$ is said to be cardinally dominated, if there exists another decision function $\underline{d} \varepsilon \tilde{D}$ such that

$$CV_i \ (\underline{d}) \geq CV_i \ (\underline{e}) \quad \text{for all i;}$$

$$CV_i \ (\underline{d}) > CV_i \ (\underline{e}) \quad \text{for all least one i.}$$

From lemma 1 it follows that an ordinarily dominated decision function is also cardinally dominated, but the converse need not be true. If a decision

function is either ordinally or cardinally dominated, it is simply called dominated.

Definition: Admissibility. A decision function $\underline{d} \epsilon D$ is said to be admissible if it is not dominated.

Note that all definitions require strict reference to the set of decision functions D under consideration.

In our example, the decision function

$$\underline{e} = \begin{bmatrix} 1, 1 \\ 1, 1 \end{bmatrix}$$ with conditional value vector (1, 1)

is ordinally dominated, for example, by

$$\underline{d} = \begin{bmatrix} 3, 5 \\ 3, 5 \end{bmatrix}$$ with conditional value vector (3, 5).

The decision functions

$$\underline{e}_1 = \begin{bmatrix} 6, 4 \\ 6, 4 \end{bmatrix}$$ with conditional value vector (6, 4) and

$$\underline{e}_2 = \begin{bmatrix} 2, 6 \\ 3, 5 \end{bmatrix}$$ with conditional value vector (8/3, 17/3)

are not ordinally dominated, but they are cardinally dominated. In fact, only the following five decision functions are admissible (not cardinally dominated):

$$\underline{d}_1 = \begin{bmatrix} 2, 6 \\ 2, 6 \end{bmatrix} \quad ; \quad \underline{d}_2 = \begin{bmatrix} 2, 6 \\ 6, 4 \end{bmatrix} \quad ; \quad \underline{d}_3 = \begin{bmatrix} 2, 6 \\ 7, 3 \end{bmatrix} \quad ;$$

$$\underline{d}_4 = \begin{bmatrix} 6, 4 \\ 7, 3 \end{bmatrix} \quad ' \quad \underline{d}_5 = \begin{bmatrix} 7, 3 \\ 7, 3 \end{bmatrix}$$

which can be easily inferred from Figure 3.

It is a well known result in statistical decision theory that in selecting decisions or decision functions that maximize expected value, the decision

maker can restrict himself to admissible gambles or decision functions. (See

for example, Ferguson, 1967; DeGroot, 1970.) Furthermore, the following

result by Blackwell and Girshick (1954) allows us to restrict our attention

to decision functions only, when selecting among gambles and decision functions:

Lemma 2 Let $\underline{d}$ be an admissible decision function with $\underline{d}(x_j) \epsilon G$ for all

j. Then $\underline{d}$ cannot be dominated by a gamble $\underline{g} \epsilon G$. The proof follows immediately

from lemma 1 which established the fact that $\underline{G} \subseteq \tilde{G} \subseteq \tilde{D}$. These implications

can be checked in our example in Figures 2 and 3.

Practically this means that a decision maker can always achieve at least

as high an expected value by first observing a free observation X and then

selecting his final gamble according to some admissible decision function

as he can by choosing among admissible gambles directly without observing X.

Consequently, we will hereafter discuss decision functions only and treat

gambles as a special case of decision functions in which $\underline{d}(x_j)$ is a constant

function. As a generic term for decision functions or gambles, we will from

now on use the term decision.

Losses Caused by Choosing Inadmissible Decisions. This section will give

a short summary of our flat maximum arguments (v. Winterfeldt and Edwards,

1973b), and we will show that these arguments do not hold for dominated de-

cisions. In our original treatment of flat maxima we asserted that if

1. S is finite;

2. A is bounded; and

3. D consists of admissible decisions only,

then the losses the decision maker will incur by selecting a suboptimal de-

cision or by using incorrect model parameters will typically be small. Basi-

cally our argument was this. If all decisions are admissible, then each will

maximize expected value with respect to some prior distribution f over the

states of nature (see Ferguson, 1967). Or, to put it in simpler terms, admissible decisions are potential candidates for optimal decisions. Conversely, for each prior distribution f over S there exists at least one admissible decision that maximizes expected value with respect to that distribution (see Ferguson, 1967). For a prior distribution f we defined a function EV*(f) as the maximum attainable expected value for that prior distribution. By the property of the expected value maximization model this function EV* will be convex and by assumption it will be bounded. All losses a decision maker can incur in a decision problem are then defined as differences between this convex and bounded EV* function and its supporting hyperplanes. From the convexity and boundedness of EV* we concluded that these losses are severely restricted in the area of an optimal decision or true parameter. These restrictions typically mean rather flat expected value functions around the optimum, as we have demonstrated in numerous examples (see v. Winterfeldt and Edwards, 1973a).

What, however, will happen, if the decision maker selects a dominated decision? Let $\underline{e}$ be the dominated decision, f be the prior distribution over the states of nature, $\underline{d}$ be an admissible decision in $\tilde{D}$ that dominates $\underline{e}$. Let, further re $\underline{d}_f$ be the optimal decision among the admissible ones. (Note that $\underline{d}_f$ does not necessarily dominate $\underline{e}$). In this case the expected loss will be

$$EL(\underline{e},f) = \sum_{i=1}^{n} f(S_i) [CV_i(\underline{d}_f) - CV_i(\underline{e})] \geq \sum_{i=1}^{n} f(S_i)[CV_i(\underline{d}) - CV_i(\underline{e})] \qquad (5)$$

in which by dominance

$$CV_i(\underline{d}) \geq CV_i(\underline{e}) \text{ for all } i \; ;$$

$$CV_i(\underline{d}) > CV_i(\underline{e}) \text{ for some } i.$$

There are, of course, no restrictions on the form of these losses within the boundaries of A.

These arguments can be demonstrated in our example problem. Figure 4 shows the expected value of the five admissible decision functions as a function of $f(S_1)$. Losses due to the selection of a suboptimal, but admissible decision function are typically small, as long as the decision functions are adjacent (in the sense that their corresponding EV-maximizing prior distributions do not differ substantially). Also, we plotted the expected value for two dominated decision functions:

$$\underline{\underline{e}}_1 = \begin{bmatrix} 1, 1 \\ 1, 1 \end{bmatrix} \quad \text{and} \quad \underline{\underline{e}}_2 = \begin{bmatrix} 1, 1 \\ 2, 6 \end{bmatrix} .$$

Figure 4 shows quite clearly that losses due to the selection of a dominated strategy can be quite substantial, regardless of the prior distribution.

Insert Figure 4 about here

We will now show that we can separate out two components of (5), one that can be attributed to the fact that $\underline{e}$ is dominated, and one that results from the suboptimality of an admissible decision, which is - in some sense - equivalent to the dominated decision $\underline{\underline{e}}$.

Definition: Admissible equivalent. An admissible equivalent of a dominated decision $\underline{\underline{e}} \epsilon D$ is an admissible decision $\hat{\underline{\underline{e}}} \epsilon \tilde{D}$ such that

$$CV_i(\hat{\underline{\underline{e}}}) = CV_i(\underline{\underline{e}}) + c \qquad \text{for all } i.$$

The admissible equivalent is determined by translating the conditional values of the dominated decision by a constant amount c, such that the translated

conditional values match those of an admissible decision. If an admissible
equivalent exists, c must be unique, since neither smaller nor larger values
could be conditional values of an admissible decision. Under some well known
conditions ($\tilde{\underline{D}}$ is closed, bounded, and convex), it can be proven that for
each dominated decision $\underline{e}$ in D there exists an admissible equivalent $\hat{\underline{e}}$ in $\tilde{D}$.
This condition will be satisfied, whenever A and S are finite (see Ferguson,
1967).

Another interpretation of an admissible equivalent $\hat{\underline{e}}$ is that its hyper-
plane as a function of f, which is defined by

$$EV(\hat{\underline{e}}, f) = \sum_{i=1}^{n} f(S_i) \; CV_i \; (\hat{\underline{e}}) \tag{7}$$

is parallel to the hyperplane of $\underline{e}$ defined by

$$EV(\underline{e}, f) = \sum_{i=1}^{n} f(S_i) \; CV_i \; (\underline{e}) \tag{8}$$

with the former hyperplane being some tangent hyperplane to EV*.

Now assume again that f is the correct prior distribution, $\underline{d}_f$ is the
optimal decision, but that the decision maker selects the dominated decision
$\underline{e}$. His loss, according to (5) will be

$$EL(\underline{e}, f) = \sum_{i=1}^{n} f(S_i) \; [CV_i(\underline{d}_f) - CV_i(\underline{e})] \tag{9}$$

which can be partitioned as follows:

$$EL(\underline{e}, f) = \sum_{i=1}^{n} f(S_i) \; [CV_i(\underline{d}_f) - CV_i(\hat{\underline{e}}) + CV_i \; (\hat{\underline{e}}) - CV_i(\underline{e})]$$

$$= \sum_{i=1}^{n} f(S_i) \; [CV_i(\hat{\underline{e}}) - CV_i(\underline{e})] + \sum_{i=1}^{n} f(S_i)[CV_i(\underline{d}_f) - CV_i(\hat{\underline{e}})]$$

$$= c + \sum_{i=1}^{n} f(S_i)[CV_i(\underline{d}_f) - CV_i(\hat{\underline{e}})] \tag{10}$$

The first part of this loss can be attributed to the fact that $\underline{e}$ is dominated and that the decision maker did not select the admissible equivalent $\hat{\underline{e}}$. The last part indicates the additional loss due to the suboptimality of the admissible equivalent $\hat{\underline{e}}$. There are, of course, no restrictions on the first part of this loss, except for the boundedness of A, but the second part is again subject to the flat maximum property.

Figure 5 demonstrates the notion of an admissible equivalent and the partition in expected losses in the example decision problem. The admissible equivalent of the decision function $\underline{e} = \begin{bmatrix} 1, & 1 \\ 2, & 6 \end{bmatrix}$ must be a randomized decision function from $\tilde{D}$, since no pure decision function has an expected value function (as a function of f) which is parallel to that of $\underline{e}$. The graph shows, how dominance (c), and suboptimality among the admissible decision functions (Δ) sum to the total loss. It also highlights the fact that the loss due to dominance in the definition of admissible equivalents is independent of the prior distribution.

Insert Figure 5 about here

We could, of course, partition the loss due to the selection of a dominated strategy in other ways than through the use of an admissible equivalent. In fact, we will do so, whenever $\underline{e}$ itself is admissible in a subset $E \subset D$. But in the absence of such a reference set E, the partition in (10) is not only plausible, but also convenient, since the loss due to dominance, c, is independent of the prior distribution f.

We will now turn to another possible partition of (5):

<u>Definition: Optimal equivalent</u>. Let $\underline{E}$ be a subset of $\underline{D}$. Let $\underline{e}_f$ be admissible in E but dominated in D. Let f be the prior distribution such that $\underline{e}_f$ is optimal in E. An optimal equivalent of $\underline{e}_f \in E$ is a decision $\underline{d} \in D$ which is optimal in D with respect to f. (Note that the optimal equivalent need not be unique.)

Assume now that the decision maker has decisions D available to him and that the prior distribution is f. Assume further that he selects a decision $\underline{e}_g$, which is dominated in D and admissible but suboptimal in E. Let $\underline{e}_f$ be the optimal decision in E with respect of f, and let $\underline{d}_f$ be its optimal equivalent. By selecting $\underline{e}_g$ instead of $\underline{d}_f$, the decision maker made two mistakes: First, restricting himself to the admissible decisions in E, he chooses a suboptimal one, and second, even the optimal one in E would be dominated by its optimal equivalent in D. His actual expected loss can be partitioned accordingly:

$$EL\ (\underline{e}_g,\ f) = EL\ (\underline{e}_f,\ \underline{d}_f) + EL\ (\underline{e}_g,\ \underline{e}_f) =$$

$$= \{\ EV_E^*\ (f) - EV\ (\underline{e}_g,\ f)\ \} +$$

$$+ \{\ EV_D^*\ (f) - EV_E^*\ (f)\ \}. \tag{11}$$

Note that EV* is defined in its restrictions to E and D, respectively here. Just as in (10), the first loss reflects suboptimality among some admissible set, which is subject to the flat maximum property; and the second loss reflects dominance and is unrestricted.

Figure 6 illustrates the concepts of an optimal equivalent and the partition in expected losses resulting from this definition using two hypothetical

EV*- functions for a two state, infinite act decision problem. The losses are indicated for the case in which the true prior probability $f(S_1) = .3$, but the decision maker selects a decision function which would be optimal among the dominated (but in themself admissible ones) under a prior of .5.

Insert Figure 6 about here

Inefficient Information Use and Dominance

We can now link the concepts of dominance and admissibility to the concept of efficiency of information use. We will show that for three plausible definitions of efficiency of information use, less efficient information use leads to decision functions that are dominated by decision functions based on more efficient information use. This means that a decision maker who processes information inefficiently, or ignores it altogether, selects dominated decision functions. Using the partitions defined in the previous section, we can then argue that some losses in a decision analysis can be attributed to inefficient use of information, and some losses can be attributed to suboptimal decisions among admissible ones. While inefficient use of information can lead - via dominance - to quite substantial losses, suboptimal admissible decision strategies will typically result only in small expected losses.

Our first definition allows us to compare information sources only if one source is completely redundant with respect to the knowledge about the states of nature given knowledge from the other source. But whenever information sources can be compared in terms of this definition, the results are very general and totally independent of the payoff structure.

Definition: Redundant information source. An information source X is said to be at least as informative as an information source Y if and only if

$$h(Y|S,X) = g(Y|X)$$

This definition and the resulting ordering of information sources in terms of their efficiency is discussed in Marschak and Radner (1972) following the more general treatment in Blackwell and Girshick (1954).

Let D be the set of decision functions based on X, and let E be the set of decision functions based on Y. Blackwell and Girshick (1954, THM 12.222) prove that, if X is at least as informative as Y, then $\tilde{E} \subset \tilde{D}$. Marschak and Radner (1972) give an informal proof that if X is at least as informative as Y, then for any prior distribution f the optimal decision function based on X is at least as valuable (in terms of EV) as the optimal decision function based on Y. From both theorems the following theorem follows immediately:

Theorem 1: If X is at least as informative as Y, then for any decision function $\underline{e}$ in E, there exists a decision function $\underline{d}$ in D, that either has the same conditional value vector as $\underline{e}$ or dominates $\underline{e}$.

The notion of a redundant information source does not lead to an ordering of all information sources, but merely to a partial ordering. Whenever two information sources can be compared in this way, however, we can reach strong conclusions about the resulting expected losses, without assuming any particular payoff structure. The main result is, of course, that admissible decision functions based on the less informative information will at best be equivalent and will usually be dominated when compared with admissible decision functions based on the more informative information source. Following

the concepts in the previous section we can then partition any losses encountered in a particular decision problem into an (unrestricted) part due to inefficiency of information (=dominance) and a restricted part due to suboptimal decision making.

Of the many examples of less informative information, the most common arise if the less informative information differs from the more informative by deletion of content only. This can occur, for example, in a situation in which X is a direct report about an event and Y is a report of the report (assuming that the originator of Y does not have interpretive information to add to the report itself). Another common class of example arises from degradation of the informational content of a technical sensor (e.g., blurring a photograph, adding white noise to a signal, etc.). Probably the most banal but common examples are those in which information is simply ignored--as is often advocated by management scientists concerned about human capacity limitations in information processing.

Our original example on p. 14 can serve as a numerical illustration of the relative effects of less informative information sources. Assume that before selecting a gamble from the set specified on p. 14 the decision maker can observe a random variable Z with the following conditional distributions:

$$g(z_1|S_1) = \frac{1}{10} \quad ; \quad g(z_2|S_1) = \frac{9}{10}$$

$$g(z_1|S_2) = \frac{9}{10} \quad ; \quad g(z_2|S_2) = \frac{1}{10} \, .$$

Assume further that instead of observing Z directly, the decision maker receives an (unreliable) report X about Z with the following distributions:

$$h(x_1|S_1,z_1) = \frac{17}{24} \quad ; \quad h(x_1|S_1,z_2) = \frac{7}{24}$$

$$h(x_2|S_1,z_1) = \frac{7}{24} \quad ; \quad h(x_2|S_1,z_2) = \frac{17}{24}$$

$$h(x_1|S_2,z_1) = \frac{17}{24} \quad ; \quad h(x_1|S_2,z_2) = \frac{7}{24}$$

$$h(x_2|S_2,z_1) = \frac{7}{24} \quad ; \quad h(x_2|S_2,z_2) = \frac{17}{24} \quad .$$

These conditional distributions of X are independent of S, thus the effect of X is simply to blur the information contained in Z. Formally Z is at least as informative as X since

$$h(X|S,Z) = h(X|Z).$$

The information impact of X with respect to knowledge about S can be inferred from these distributions as:

$$h(x_1|S_1) = \frac{1}{3} \quad ; \quad h(x_2|S_1) = \frac{2}{3}$$

$$h(x_1|S_2) = \frac{2}{3} \quad ; \quad h(x_1|S_2) = \frac{1}{3} \quad .$$

Note that these are exactly the conditional distributions from our original example on p. 15.

Figure 7 plots the EV*-functions for decision functions D based on Z and for decision functions E based on X. The dotted line $EV*_p(f)$ represents the EV*-function for perfect information. $EV(\underline{e},f)$ is the EV-function for the decision function $\begin{bmatrix} 2 & 6 \\ 6 & 4 \end{bmatrix}$, which is admissible in E but dominated in D. c and Δ indicate the losses the decision maker will incur if the correct prior probability is .5 and he uses X as his information source (c), and $\underline{e}$ as his

decision function (Λ). As long as the prior distributions are not extreme, even gross misrepresentations of his prior opinion will cost the decision maker less than the loss in information.

Insert Figure 7 about here

Especially subtle and important are cases in which the human intuitive processing of the information, rather than its raw content, renders the information less informative. This can obviously happen if that processing ignores part of the information. Some interpretations of the well-established behavioral phenomenon of conservatism in human probabilistic information processing (see for example Edwards, 1968; Slovic and Lichtenstein, 1971) would include it within this category.

It seems clear that Fryback's (1974) data, which originally started us thinking about this sort of possibility, fit this case. Fryback found essentially two sources of major suboptimality. One was a clearly perverse strategy: That of performing a costly diagnostic procedure to "rule out" an unlikely but anxiety-producing hypothesis, when the alternative is to perform a less costly procedure that will in any case be necessary if the more likely, less drastic hypothesis is true and that will, if positive, confirm that hypothesis. The other was that among his respondents, all expert at reading the kind of x-rays he was studying, one was such a super-expert that, even though he used a fairly poor strategy in a decision-theoretical sense, he could do very much better than anyone else simply because of his radiological expertise.

The latter source of suboptimality directly fits this case. The former does not; it is an example of the kind of error caused by insufficiently deep

deliberation that we ruled out of consideration at the beginning of this paper. Yet even here, better diagnosis, by making the anxiety-producing hypothesis exceedingly unlikely instead of merely unlikely, might have reduced the incidence, and hence the expected cost, of perverse error.

Our second definition can only be applied in relatively special decision problems, but it does not require such strict redundancy as did our first. We assume a two states, two acts decision problem, in which the decision maker can choose between two gambles

$$\underline{g}_1 = (a, b)$$
$$\underline{g}_2 = (c, d)$$

The only restrictions on the outcomes are that $a > b$ and $c < d$. Before making his decision, a decision maker can observe a value of a random variable X, with distribution depending on the states of nature. Finally, we make the assumption of a monotone likelihood ratio; that is,

$$\frac{g(x|S_2)}{g(x|S_1)} \geq \frac{g(x'|S_2)}{g(x'|S_1)} \qquad \text{iff} \qquad x \geq x' \; .$$

Definition: More sensitive information source. Let the above assumptions be true for two information sources X and Y. Let $F_X(X|S_i)$ and $F_Y(Y|S_i)$ be the conditional cumulative probability distributions of X and Y, respectively. X is said to be more sensitive than Y if there exists a transformation $Y' = Y+C$ such that

1) $F_{Y'}(Y'=z|S_1) < F_X(X=z|S_1)$ for all z, and

2) $F_{Y'}(Y'=z|S_2) > F_X(X=z|S_2)$ for all z.

These two conditions are equivalent to saying that, given S_1, F_X stochastically dominates $F_{Y'}$, and, given S_2, $F_{Y'}$ stochastically dominates F_X. For a definition of stochastic dominance (not to be confused with the concepts of ordinal and cardinal dominance defined above) see Lehman, 1959.

Loosely speaking this definition means that the two distributions of X are more separated than the two distributions of Y. More precisely, this will be the case whenever Y can be translated so that its two cumulative probability distributions lie totally between the two cumulative probability distributions of X.

<u>Theorem 2</u>: Under the above assumptions, let X be more sensitive than Y. Let D be the set of decision functions based on X and let E be the set of decision functions based on Y. Then for any admissible decision function $\underline{e} \epsilon E$, there exists an admissible decision function $\underline{d} \epsilon D$ that dominates $\underline{e}$.

Proof: Consider Y first. From the Neyman-Pearson lemma (see Lehman, 1959) and from the monotonicity of the likelihood ratio it follows that any admissible decision function $\underline{e} \epsilon E$ must be of the form

$$\underline{e}\,(Y) = \begin{cases} \underline{g}_1 & \text{if } Y \le x_0 \\ \\ \underline{g}_2 & \text{if } Y > x_0 \end{cases}$$

The conditional values of $\underline{e}$ are therefore

$$CV_1(\underline{e}) = a\,F_Y(x_0|S_1) + c\,[1-F_Y(x_0|S_1)]$$

$$CV_2(\underline{e}) = b\,F_Y(x_0|S_1) + d\,[1-F_Y(x_0|S_1)].$$

Let C be such that the conditions of definition 2 are fulfilled. Consequently

$$F_{Y'}(x_0 + C|S_i) = F_Y(x_0|S_i).$$

Now consider X. Again admissible decision functions $\underline{d} \epsilon D$ must be of the form

$$\underline{d}(X) = \begin{cases} \underline{g}_1 & \text{if } X \leq x_0' \\ \\ \underline{g}_2 & \text{if } X > x_0' \end{cases}$$

Let $x_0' = x_0 + C$. Then the conditional values of $\underline{d}$ are

$$CV_1(\underline{d}) = a \quad F_X(x_0 + C|S_1) + c[1 - F_X(x_0 + C|S_1)]$$

$$CV_2(\underline{d}) = b \quad F_X(x_0 + C|S_2) + d[1 - F_X(x_0 + C|S_2)].$$

To prove the theorem, all we will show is that

$$CV_i(\underline{d}) > CV_i(\underline{e}) , \quad i=1, 2.$$

By assumption

$$a > c , b < d.$$

Furthermore, by stochastic dominance (1 and 2)

$$F_{Y'}(x_0 + C|S_1) = F_Y(x_0|S_1) < F_X(x_0 + C|S_1) \text{ and}$$

$$F_{Y'}(x_0 + C|S_2) = F_Y(x_0|S_2) > F_X(x_0 + C|S_2).$$

Therefore

$$a \quad F_X(x_0 + C|S_1) + c[1 - F_X(x_0 + C|S_1)] >$$

a $\quad F_Y(x_0|S_1) + c[1-F_Y(x_0|S_1)] \quad$ and

b $\quad F_X(x_0 + C|S_2) + d[1-F_X(x_0 + C|S_2)] >$

b $\quad F_Y(x_0|S_2) + d[1-F_Y(x_0|S_2)]$

which establishes the fact that for any admissible decision function $\underline{e}\epsilon E$ (with corresponding cutoff value x_0) there exists a decision function $\underline{d}\epsilon D$ (with corresponding cutoff value $x_0 + C$) that dominates $\underline{e}$.

The following military decision problem illustrates the concepts of sensitivity of observations. A commander of a battleship observes an unidentified ship moving towards his own ship. Attempts to establish radio contact with the oncoming ship fail. He has to decide whether or not to attack, considering that the ship may be either an enemy (S_1) or an ally (S_2). Assume that the only information which differentiates between the two states of the ship is its length and that an enemy ship is longer than an allied ship. The commander can obtain a length estimate from his own position (Y) or from a nearby ally ship which has a better angle and is closer to the unidentified ship (X). Assuming normal measurement errors, let

$$F_Y(.|S_i) \sim N(m_i, s) \quad \text{and}$$

$$F_X(.|S_i) \sim N(m_i', s),$$ and assume that the better position of the allied observer is expressed in the fact that $m_1'-m_2' > m_1-m_2$. We can now show that X is more sensitive than Y.

Since

$$0 < m_1 - m_2 < m_1' - m_2' \quad,$$

we can find a C such that

$$m_2' < m_2 + C < m_1 + C < m_1'.$$

Let $Y' = Y+C$. Since all distributions involved are normal differing only in mean, it immediately follows that

$$F_X (X = z|S_1) > F_{Y'} (Y' = z|S_1) \text{ for all } z, \quad \text{and}$$

$$F_X (X = z|S_2) < F_{Y'} (Y' = z|S_2) \text{ for all } z.$$

Also, since all distributions are normal, with equal variance, the likelihood ratio is monotone. Therefore, all conditions of our definition are fulfilled, and X is more sensitive than Y.

The conclusion that decision functions based on X will dominate decision functions based on Y can be demonstrated with the following payoff matrix:

$$
\begin{array}{ccc}
 & \text{Enemy} & \text{Ally} \\
\text{attack} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
\text{do not attack} &
\end{array}
$$

Further specifications are that $s=1$, and that $m_1 - m_2 = d_Y' = 1$, and that $m_1' - m_2' = d_X' = 2$. Figure 7 plots the conditional values and demonstrates that admissible decision functions based on Y are dominated by admissible decision functions based on X. Figure 8 shows the corresponding EV*-functions $EV*_E$ and $EV*_D$ together with some potential losses due to dominance and sub-optimality among admissible decision functions.

Insert Figures 7 and 8 about here

Our final definition of the efficiency of information is motivated by its ability to reduce the variance of the posterior probability distribution over the states of nature. We have to assume that state and act spaces are real valued, and that the value function is quadratic in both. Specifically, if the state of nature is s and the decision maker selects g, then his outcome will be $K - (s-g)^2$.

Definition: Precision of information. Information X is said to be more precise than information Y if and only if

$$E\{VAR[S|X]\} = \sum_{j=1}^{m} g(x_j) VAR [S|x_j] < \sum_{r=1}^{\ell} h(y_r) VAR[S|y_r] = E\{VAR[S|Y\} \qquad (13)$$

where $VAR(S|x_j)$ is the variance of the (posterior) distribution, if $X=x_j$.

Let $\underline{e}$ be a decision function based on Y and let $\underline{d}$ be a decision function based on X. The following is a standard result in statistical decision theory (see Ferguson, 1967; DeGroot, 1970):

Lemma 4   Under the above assumptions the expected values of a decision function $\underline{e}$ and a decision function $\underline{d}$ are

$$EV(\underline{d}, f) = K - \sum_{j=1}^{m} g(x_j) VAR (S|x_j) - \sum_{j=1}^{m} g(x_j) [E(S|x_j) - \underline{d}(x_j)]^2 \qquad (14)$$

$$EV(\underline{e}, f) = K - \sum_{r=1}^{\ell} h(y_r) VAR (S|y_r) - \sum_{r=1}^{\ell} h(y_r) [E(S|y_r) - \underline{e}(y_r)]^2. \qquad (15)$$

The second term of both expected value formulas is the expected posterior variance and corresponds to our definition of precision of the information X and Y. The third term can be made equal to zero by letting

$$\underline{d}(x_j) = E(S|x_j) \qquad \text{and}$$

$$\underline{e}(y_r) = E(S|y_r).$$

The following theorem is an immediate consequence of the lemma:

Theorem 3: Assume that the value function is quadratic in the state and act variables. Assume that information X is more precise than information Y. Then for any decision function $\underline{e}$ based on Y there exists a decision function $\underline{d}$ based on X that dominates $\underline{e}$.

Proof: By assumption

$$\sum_{j=1}^{m} g(x_j) \text{ VAR } (S|x_j) < \sum_{r=1}^{\ell} h(y_r) \text{ VAR } (S|y_r).$$

Now take any $\underline{e}\epsilon E$ and choose $\underline{d}$ such that $\underline{d}(x_j) = E(S|x_j)$. Then

$$EV(\underline{d},\ f) = K - \sum_{j=1}^{m} g(x_j) \text{ VAR } (S|x_j) >$$

$$K - \sum_{r=1}^{\ell} h(y_r) \text{ VAR } (S|y_r) - \sum_{r=1}^{\ell} h(y_r) [E(S|y_r) - \underline{e}(y_r)]^2 = EV(\underline{e},\ f).$$

So that for any $\underline{e}\epsilon E$ there exists a $\underline{d}\epsilon D$ such that

$$EV(\underline{d},\ f) > EV(\underline{e},\ f)$$

independent of the prior distribution f. Since there is no prior distribution for which $\underline{e}\epsilon E$ would be optimal, it follows (see Ferguson, 1967), that $\underline{e}$ must be dominated.

As an example, consider the following simple inventory problem. The decision maker has to stock his store with a certain supply S of some good. His profit will depend on the unknown demand D. Assume that his profit for

any supply/demand situation can be expressed as

$$V(S,D) = K - (S-D)^2.$$

Before purchasing the good, the decision maker can observe a random variable X with distribution depending on the true demand D. Specifically, we assume that $g(X|D=d)$ is normal with expectation d and variance s_X^2. Another random variable Y also has a distribution depending on D which is normal with expectation d and variance s_Y^2. Assume $s_Y^2 > s_X^2$. Under these conditions, we can show that X is more precise than Y, if the prior distribution over the demand is also normal with mean $\overline{d}$ and variance s^2.

The following is a standard result for the above conjugate distributions:

$$VAR(D|X) = (1/s^2 + 1/ s_X^2)^{-1}$$

$$VAR(D|Y) = (1/s^2 + 1/ s_Y^2)^{-1}$$

Since both conditional variances are independent of the specific values of X and Y, and since $s_X^2 < s_Y^2$, we have

$$VAR(D|X) \; : \; VAR(D|Y) \quad \text{for all X and Y, and}$$

$$E\{VAR(D|X)\} < E\{VAR(D|Y)\}.$$

That is, X is more precise than Y.

The loss due to dominance of Y is $(\frac{1}{s^2} + \frac{1}{s_X^2})^{-1} - (\frac{1}{s^2} + \frac{1}{s_Y^2})^{-1}$.

Losses due to suboptimality among the admissible decision functions based on Y are quadratic with a minimum at

$$\underline{\underline{e}}(Y) = E(D|Y) = \frac{\frac{1}{s} \cdot \overline{d} + \frac{1}{s_Y} \cdot Y}{\frac{1}{s} + \frac{1}{s_Y}} \quad .$$

The proof shows us a convenient way to partition the losses into those caused by the selection of a less precise information source and those caused by an additional suboptimality of the decision function based on that information source. Let $\underline{\underline{e}}$ be a suboptimal and dominated decision function based on Y. Then the loss resulting from choosing $\underline{\underline{e}}$ instead of the optimal $\underline{d}$ is

$$EL(\underline{\underline{e}}, f) = \{ \sum_{r=1}^{\ell} h(y_r) \, VAR(D|y_r) - \sum_{j=1}^{m} g(x_j) \, VAR(D|x_j) \} + \sum_{r=1}^{\ell} h(y_r)$$

$$[E(D|y_r) - \underline{\underline{e}}(y_r)]^2 \tag{16}$$

The second part is a quadradic function with a minimum at $\underline{\underline{e}}(y_r) = E(D|y_r)$. It is subject to the flat maximum property, since it represents suboptimality among the admissible decision functions in E. The first part is caused by the inefficiency of the information variable Y relative to X. This loss is restricted only in the limits between 0 and the variance of the prior distribution.

There probably are other definitions of efficiency of information sources that lead to similar conclusions about dominated decision functions, but the three definitions used here not only illustrate the main point, but also cover a rather wide range of decision problems. The main conclusion is that for all three definitions of efficiency we find that inefficient use of information can, and normally will, lead to use of a dominated decision strategy that may cause large losses for the decision maker.

Discussion

Our message is this. If a decision problem is properly structured and optimal use has been made of the best relevant information bearing on it (taking into account if necessary the costs of doing so compared with the costs of using inferior information or none), a decision maker can be fairly sure of making a fairly good decision, though not necessarily the optimal one, even if his prior opinions have been inaccurately elicited. The mathematics behind this assertion state it as a typical but not inevitable result; they naturally say very little about any specific decision problem. Still, they suggest that formulating the problem and processing the information are the heart of the task; elicitation of probabilities (and of values, though that topic is more complicated) are secondary.

The mathematics also offer analysis of admissibility and of EV*-functions as important tools of sensitivity analysis. In a way, EV*-functions are loss-generating functions. Their shapes and differences control sensitivity in specific decision problems. Whether or not our typical (flat maximum) result holds in a specific problem can often be inferred, even without formal analysis, by determining boundary losses by means of EV*-functions; for example, by comparing losses with no information to losses with perfect information; or by imposing boundaries on the ranges of prior distributions or of admissible decision strategies.

We are just beginning to understand the problem of sensitivity in decision analysis. So far we have only looked at losses produced by

1) incorrect assessment of prior distributions;

2) suboptimal but admissible decision making;

3) inadmissible decision making;

4) inefficient information processing.

Nothing is known yet about other errors in decision analysis, such as leaving out value dimensions, not considering all available action alternatives, not specifying states of nature or information sources finely enough.

Ideally, a general approach to the sensitivity/insensitivity issue in decision analysis should provide the decision analyst with three kinds of information:

1) a rank order of the parts of a decision analysis according to their typical sensitivity/insensitivity;

2) examples that fit certain decision problems through simple parametrization;

3) specific tools for sensitivity analyses in particular decision problems.

So far our answers are still very incomplete. Generally, we think that admissibility analysis and the information processing part are more loss sensitive in decision analysis than are actual admissible decision making among admissible options, or elicitation of prior distributions. We gave some examples and some tools for sensitivity analysis--the former probably too specific, the latter definitely too general. But in spite of the incompleteness of what has been done, we think our two main results are useful both to decision analysts and to research on decision theory. For analysts, we have already suggested that the structuring and information processing parts of the analysis--the hard parts--are also the important ones; most analysts knew that already. For researchers, our message is that research on the merits of information sources, on optimization of information processing, and on formulation of decision problems is more important than work on precise elicitation and optimization procedures.

REFERENCES

Blackwell, D. and Girshick, M. A. Theory of Games and Statistical Decisions.
    New York: Wiley, 1954.

Brown, R. V.; Kahr, A. S.; and Peterson, C. Decision Analysis for the
    Manager. New York: Holt, Rinehart, and Winston, 1974.

DeGroot, M. H. Optimal Statistical Decisions. New York: McGraw-Hill, 1970.

Edwards, W.; Lindman, H.; and Savage, L. J. Bayesian statistical inference
    for psychological research. Psychological Review 1963, 70, 193-242.

Edwards, W. Conservatism in human information processing. In Kleinmuntz, B.
    (ed.) Formal Representations of Human Judgment. New York: Wiley, 1968.

Fryback, D. G. Use of radiologists' subjective probability estimates in a
    medical decision making problem. Michigan Mathematical Psychology Pro-
    gram, MMPP 74-14, The University of Michigan, Ann Arbor, 1974.

Ferguson, T. S. Mathematical Statistics: A Decision Theoretic Approach.
    New York: Academic Press, 1967.

Green, D. M. Psychoacoustics and detection theory. Journal of the Acousti-
    cal Society of America 1960, 32, 1189-1203.

Lehman, E. L. Testing Statistical Hypotheses. New York: Wiley, 1959.

Marschak, J. and Radner, R. Economic Theory of Teams. New Haven: Yale
    University Press, 1972.

Murphy, A. H. and Winkler, R. L. Scoring rules in probability assessment
    and evaluation. Acta Psychologica 1970, 34, 273-286.

Raiffa, H. Decision Analysis: Introductory Lectures on Choices under
    Uncertainty. Reading, Massachusetts: Addison Wesley, 1968.

Savage, L. J. The Foundations of Statistics. New York: Wiley, 1954.

Schlaifer, R. Analysis of Decisions under Uncertainty. New York: McGraw-
    Hill, 1969.

v. Winterfeldt, D. and Edwards, W. Costs and payoffs in perceptual research.
    Technical Report No. 011313-1-T. Engineering Psychology Laboratory,
    University of Michigan, Ann Arbor, 1973a.

v. Winterfeldt, D. and Edwards, W. Flat maxima in linear optimization
    models. Technical Report No. 011313-4-T. Engineering Psychology
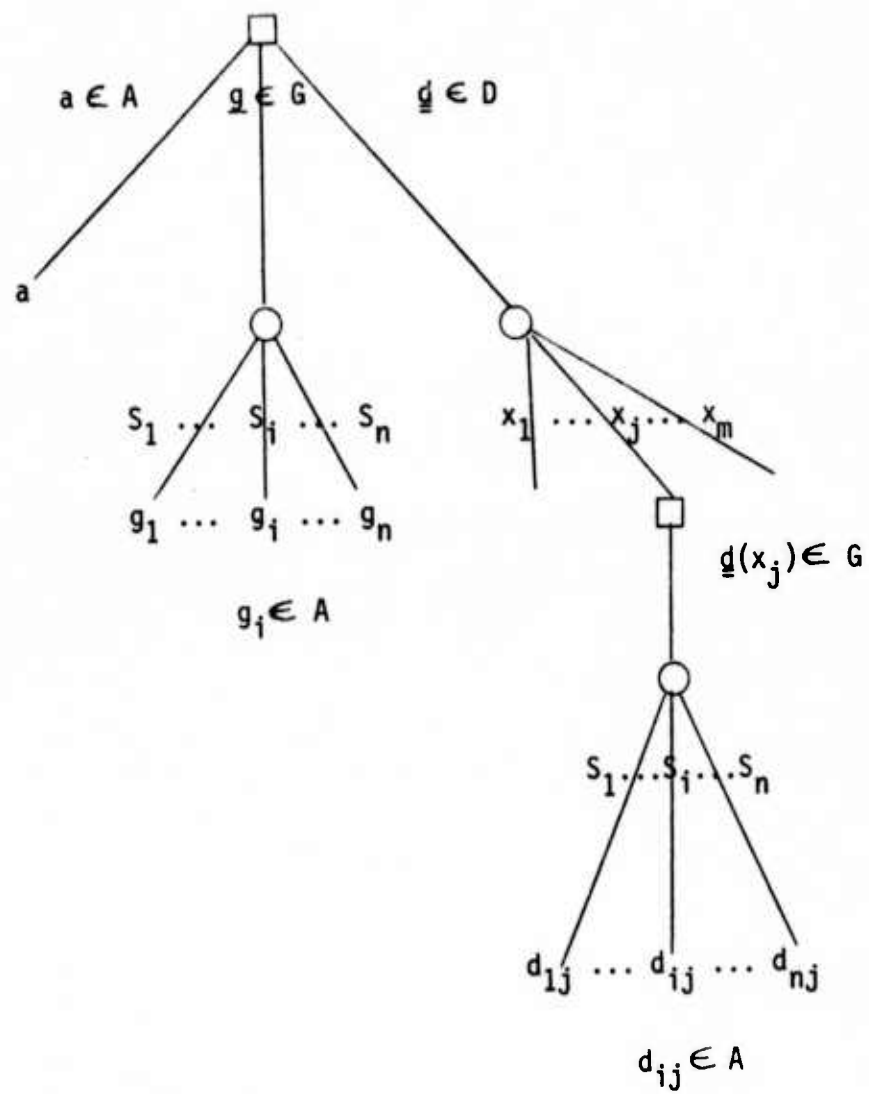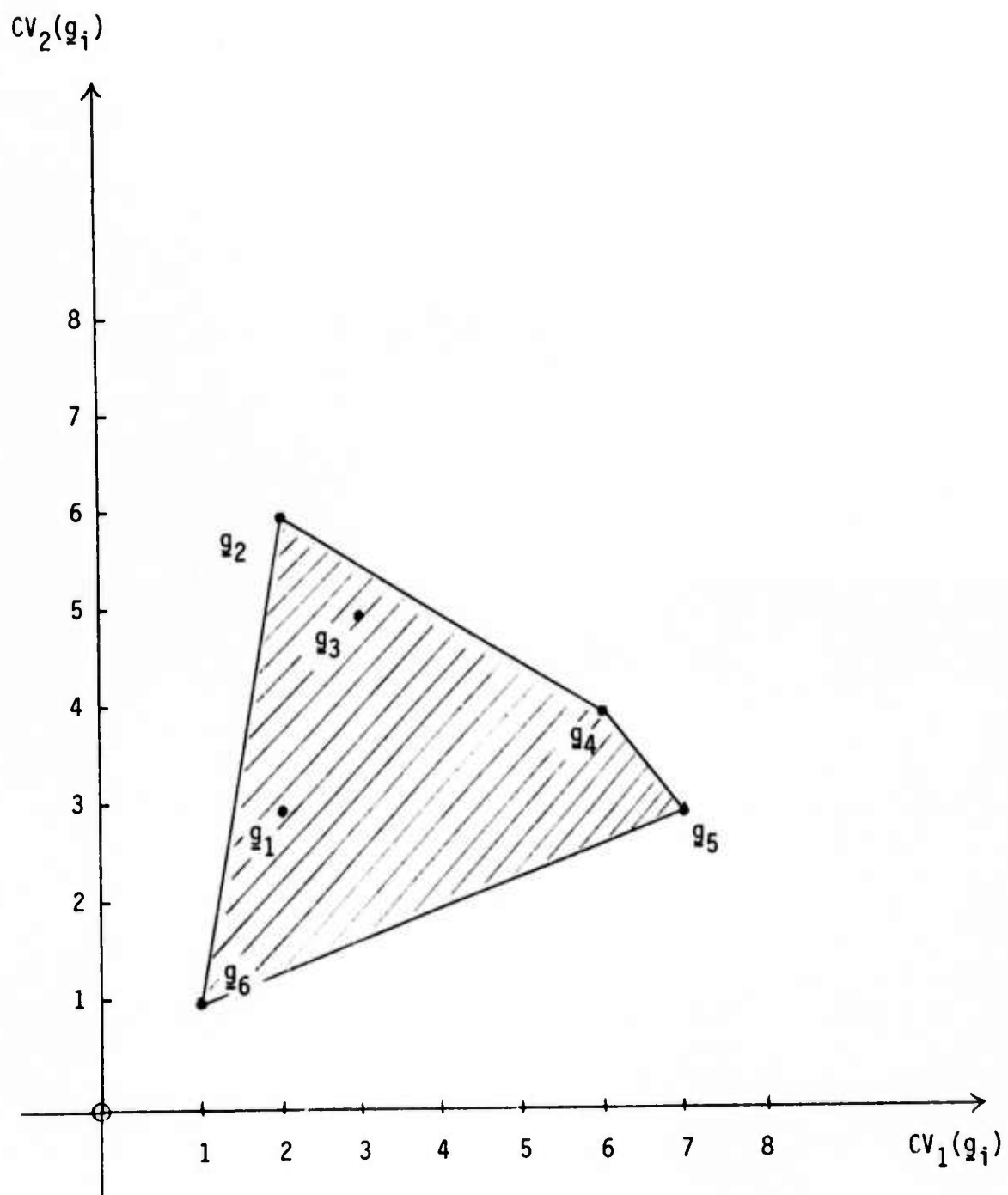   Laboratory, University of Michigan, Ann Arbor, 1973b.

LIST OF FIGURES

Figure 1

Figure 2

Figure 3

Figure 4
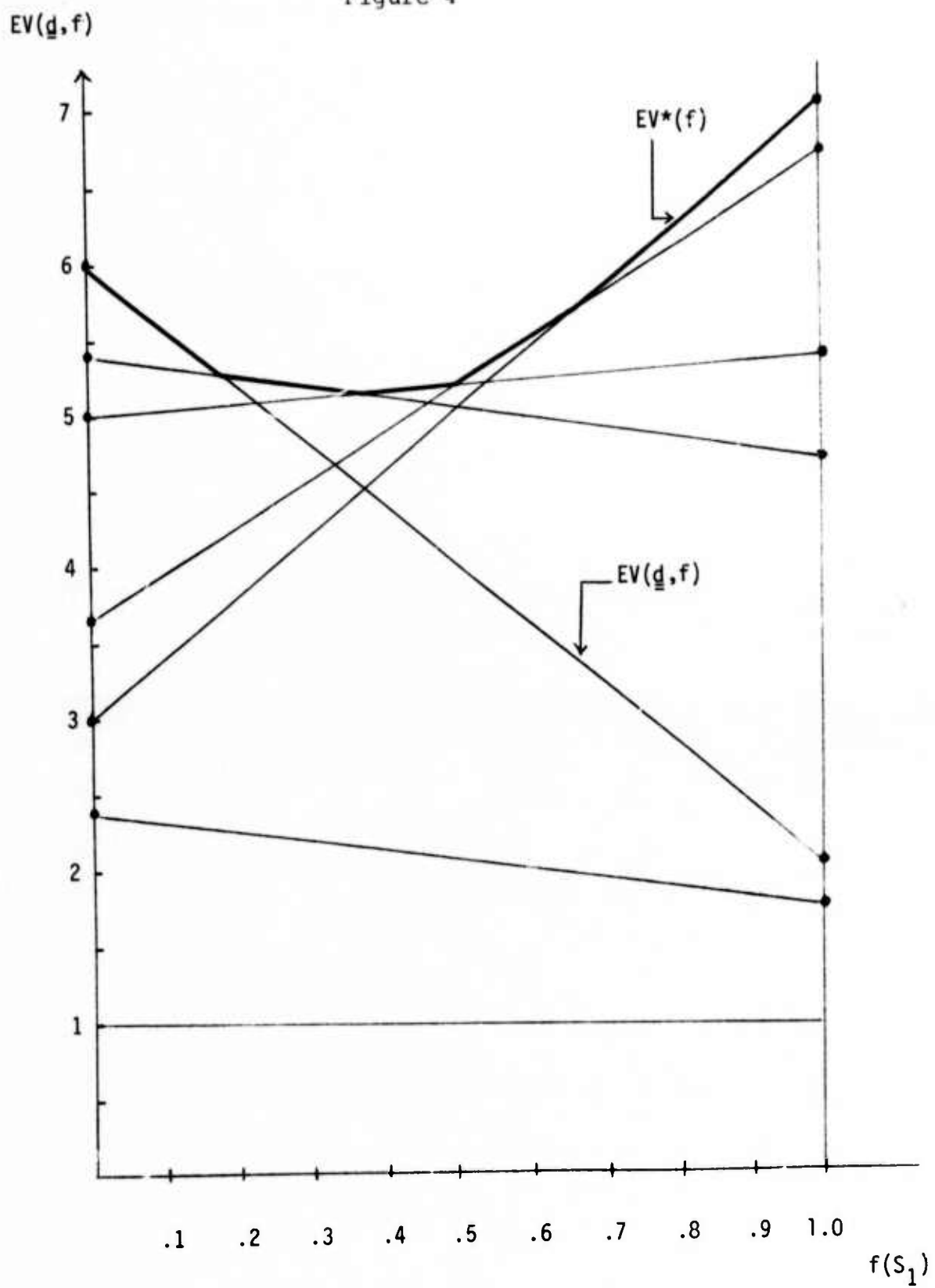
Figure 5
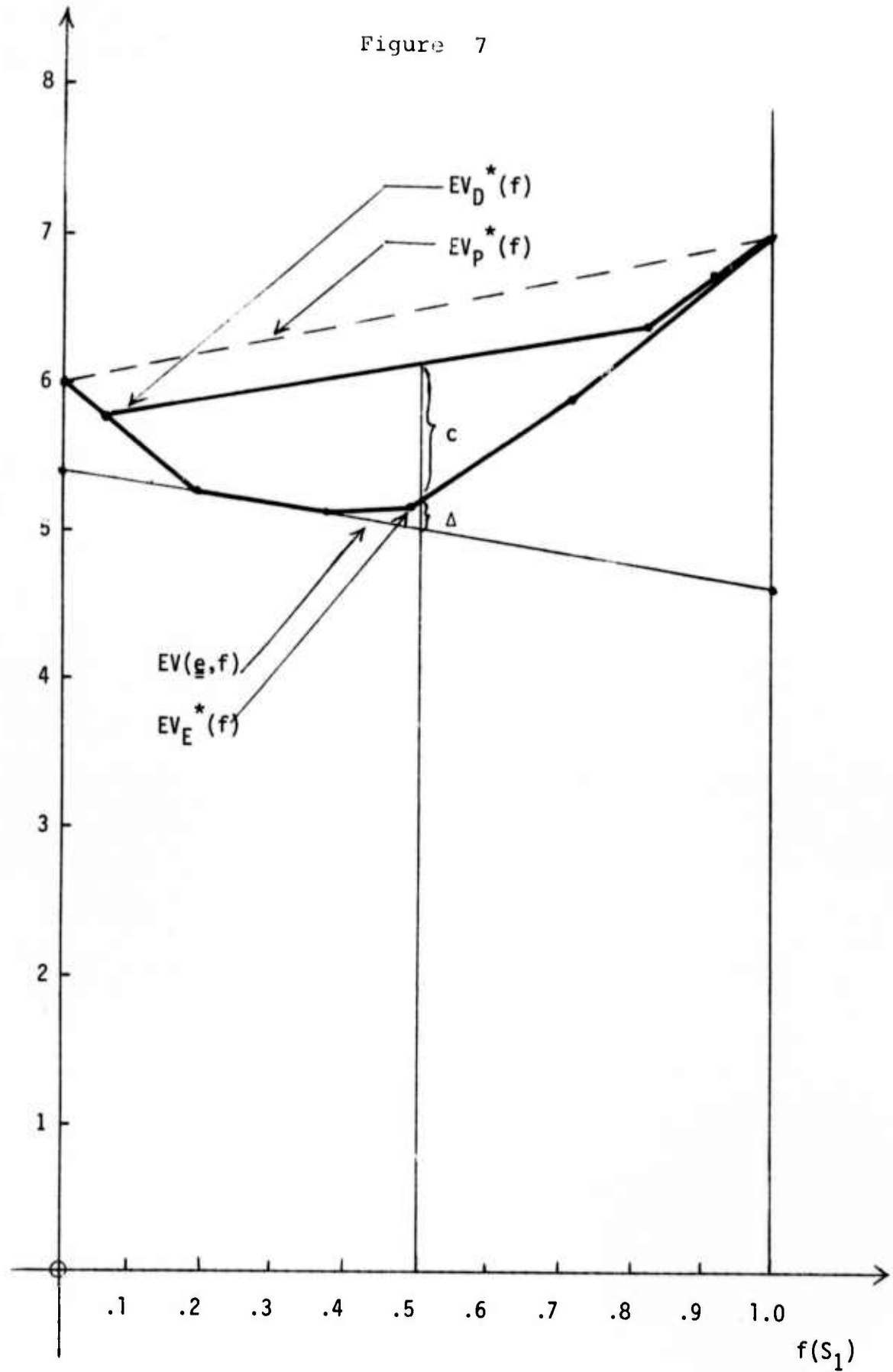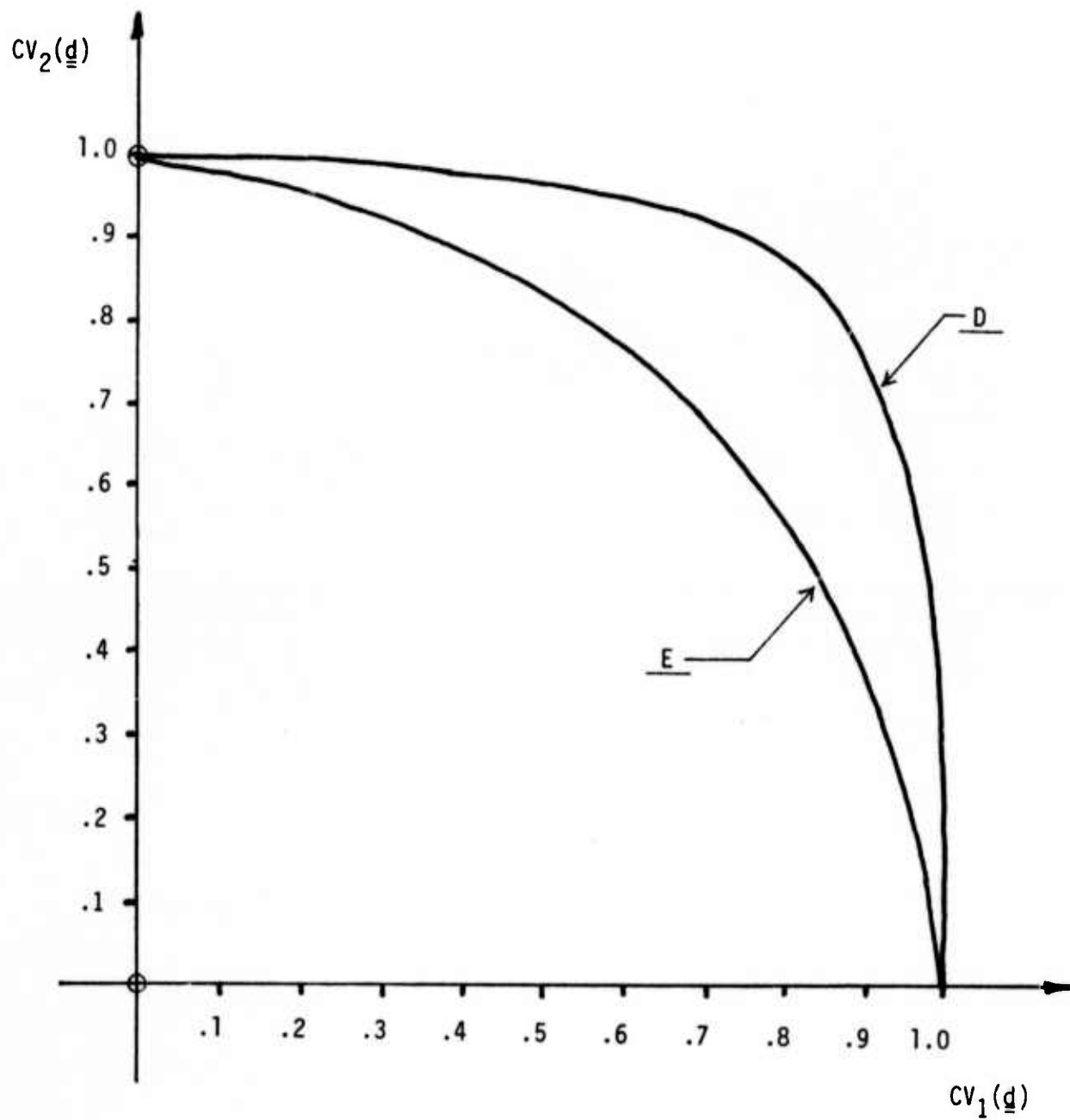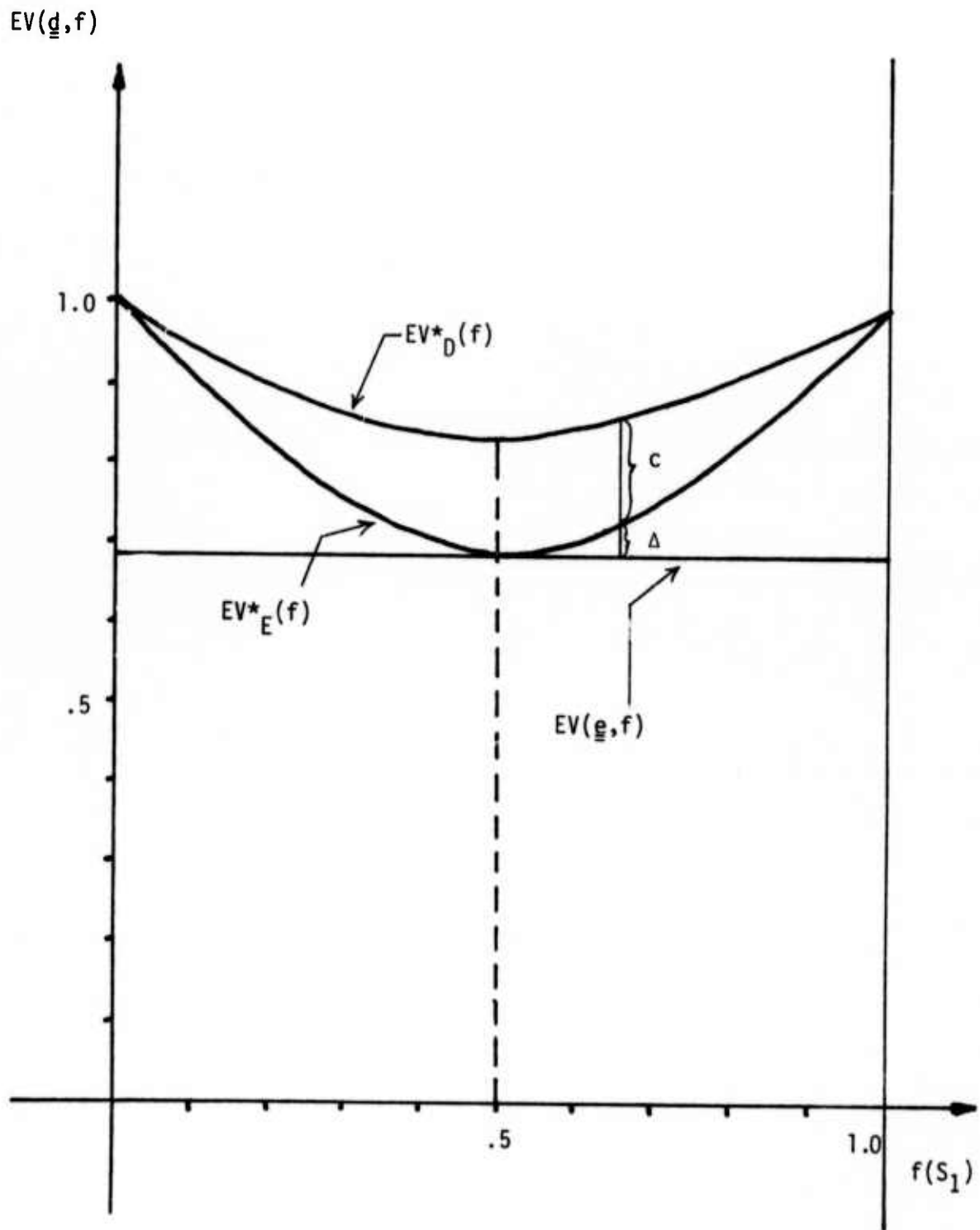
Figure 6

Figure 7

Figure 8

Figure 9

54<
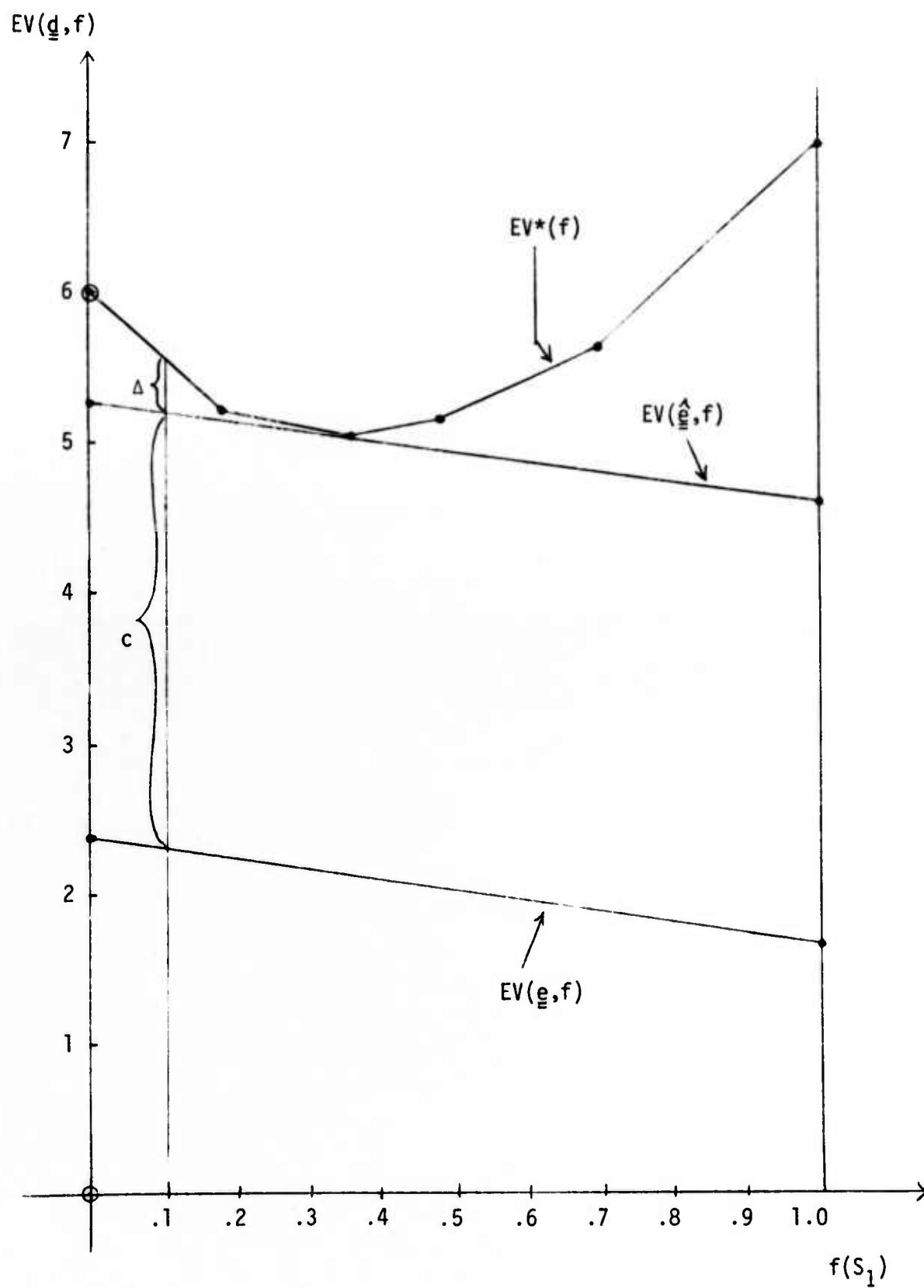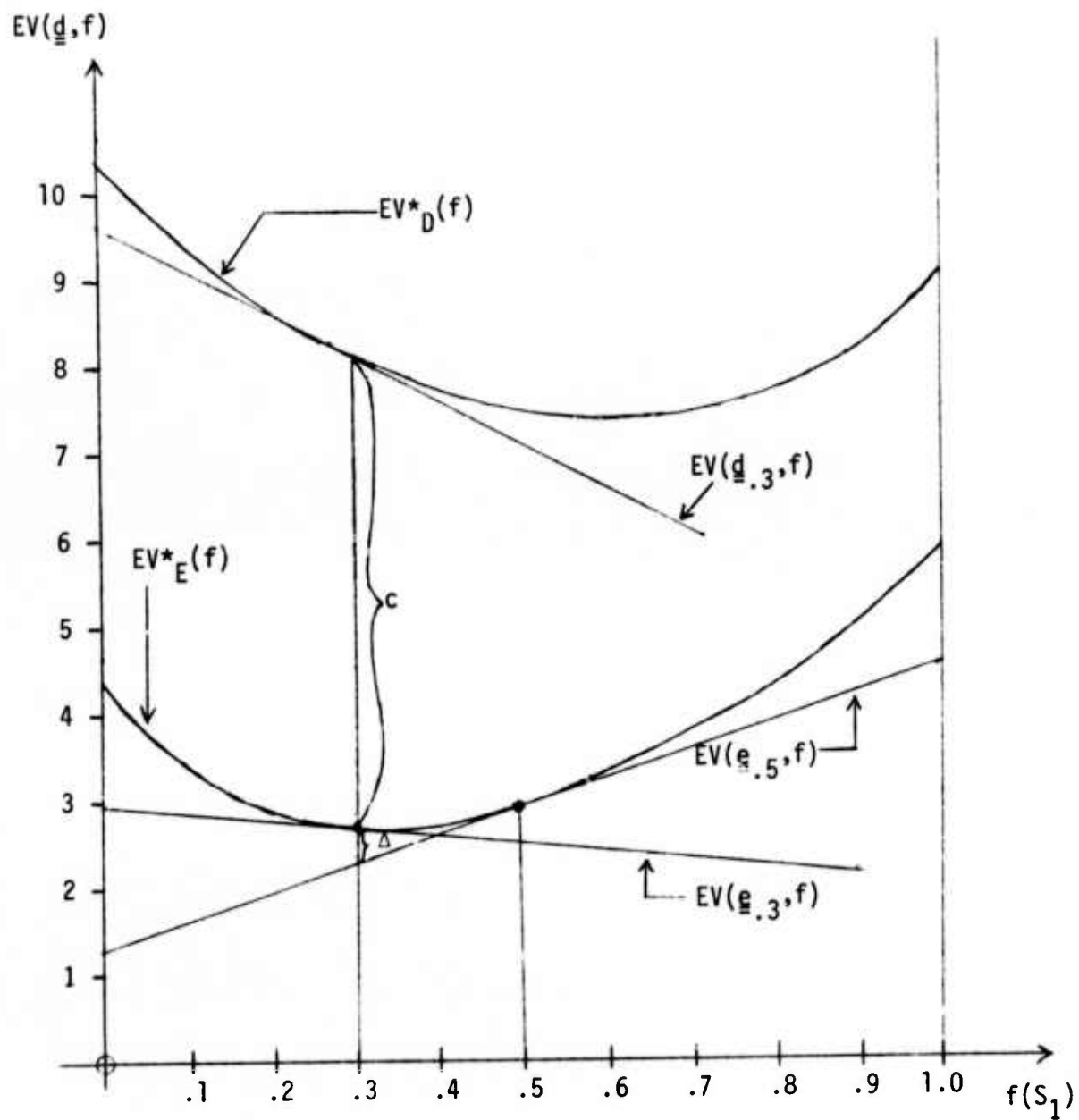
Social Science Research Institute
Research Reports

75-1   N. Miller and G. Maruyama. Ordinal Position and Peer Popularity.
January, 1975.

75-2   G. Maruyama and N. Miller. Physical Attractiveness and Classroom Acceptance.
January, 1975.

75-3   J.R. Newman, S. Kirby and A.W. McEachern. Drinking Drivers and Their Traffic Records.
February, 1975.

75-4   Detlof von Winterfeldt and Ward Edwards. Error in Decision Analysis: How to Create the
Possibility of Large Losses by Using Dominated Strategies.
April, 1975.